

The Complexity of Antidifferentiation, Denjoy Totalization, and Hyperarithmetic Reals

ALEXANDER S. KECHRIS

1. The problem of the primitive. We consider real functions on the interval $[0, 1]$. Denote by Δ the set of derivatives; i.e.,

$$\begin{aligned}\Delta &= \{f: f \text{ is a derivative}\} \\ &= \{f: \exists F: [0, 1] \rightarrow \mathbf{R} \text{ (} F \text{ is differentiable and } f = F')\}.\end{aligned}$$

If $f \in \Delta$, any F with $F' = f$ is a *primitive* of f and is uniquely determined up to a constant. To normalize, we denote by $F(x) = \int_0^x f$ the unique primitive of f with $F(0) = 0$. This is the original Newtonian concept of integration as the inverse operation of differentiation, i.e., antidifferentiation.

We will be concerned here with some definability aspects of the

CLASSICAL PROBLEM OF THE PRIMITIVE. Reconstruct the primitive F of a given derivative f .

This goes back to Newton and Leibniz and has been considered over the years in a different light as the concept of function has evolved (see [L2, P, S]). In “modern times” this problem was solved by Cauchy for continuous f (Cauchy definition of the integral) and Riemann for Riemann-integrable f (Riemann integration). Various generalizations were developed in the last half of the nineteenth century until Lebesgue introduced in his thesis (“Intégrale, Longueur, Aire,” 1902) his concept of integral. As explained in his book *Leçons sur l'Intégration et la Recherche des Fonctions Primitives* [L1], published in 1904, his primary motivation was the solution of the problem of the primitive in the general case of an arbitrary derivative. His Lebesgue integral made a major breakthrough here and totally resolved the problem in the case of a bounded derivative, more generally a Lebesgue integrable one. However, Lebesgue regretfully pointed out that he was unable to solve the problem completely for nonintegrable f . He considered this a major open problem that awaited solution. Such a solution was finally achieved in 1912 by A. Denjoy. The second edition of Lebesgue's aforementioned book [L2] published in 1928, contains an extensive treatment of Denjoy's work.

Research partially supported by National Science Foundation Grant DMS-8416349.

2. Denjoy totalization. In his solution to the problem of the primitive, Denjoy developed a concept of integration called *Denjoy totalization*. It consists of a transfinite iteration of Lebesgue integrations, computations of limits of sequences, and summations of series. (It is worth noting that transfinite iteration has also been used earlier in the context of Riemann as well as Lebesgue integration—even by Lebesgue himself.) We will briefly review now the concept of Denjoy totalization.

We begin with the following result of Lebesgue.

PROPOSITION 2.1. *Let $E \subseteq [a, b]$ be closed and $\{(a_i, b_i)\}$ the intervals contiguous to E in $[a, b]$ (i.e., the components of its complement). Let F be differentiable on $[a, b]$ and assume F' is Lebesgue integrable on E and $\sum |F(b_i) - F(a_i)| < \infty$. Then*

$$F(b) - F(a) = \int_E F'(x) dx + \sum [F(b_i) - F(a_i)].$$

This motivates the following definitions.

DEFINITION 2.2. Let E be a closed set and f a Lebesgue measurable function. We say that a point $x \in E$ is a point of *nonsummability* of f on E if f is not Lebesgue integrable in every $I \cap E$, I an open interval containing x .

DEFINITION 2.3. Let F be a continuous function on $[a, b]$, let $E \subseteq [a, b]$ be closed, and let $\{(a_i, b_i)\}$ be the contiguous intervals of E in $[a, b]$. A point $x \in E$ is called a point of *divergence* of F on E if $\sum_I |F(b_i) - F(a_i)| = \infty$ for all open intervals I containing x , where \sum_I indicates that we include only the (a_i, b_i) contained in I .

One has now the following fact.

PROPOSITION 2.4. *If F is a differentiable function on $[a, b]$ and $E \subset [a, b]$ is closed, then the points of nonsummability of F' on E and the points of divergence of F on E form a closed nowhere dense set in E .*

We can describe now the process of Denjoy totalization:

Given $f(= F')$ we reconstruct the primitive F —or actually the differences $F(b) - F(a)$ for all $a < b$ in $[0, 1]$ —by transfinite induction as follows. We construct a decreasing transfinite sequence $E_1 \supseteq E_2 \supseteq E_3 \supseteq \cdots \supseteq E_\alpha \supseteq \cdots$ of closed subsets of $[0, 1]$, each of which is nowhere dense in its predecessor (at successor stages), and simultaneously for each α the differences $F(b) - F(a)$ for (a, b) disjoint from E_α . Since for some countable α_0 we must have $E_{\alpha_0} = \emptyset$, it follows that by stage α_0 we have constructed $F(b) - F(a)$ for all $a < b$ in $[0, 1]$; i.e., we have reconstructed the primitive of f .

To start with, we let

$$E_1 = \{x \in [0, 1] : x \text{ is a point of nonsummability of } f \text{ on } [0, 1]\}.$$

Let $\{(a_i^1, b_i^1)\}$ be the contiguous intervals of E_1 in $[0, 1]$. Then $F(b) - F(a)$ is computed by Lebesgue integration for all $a < b$ with $[a, b] \cap E_1 = \emptyset$, and thus by taking limits (since F is continuous) we can compute $F(b_i^1) - F(a_i^1)$ as well.

Now let

$$E_2 = \{x \in E_1 : x \text{ is a point of nonsummability of } f \text{ on } E_1 \\ \text{or } x \text{ is a point of divergence of } F \text{ on } E_1\}.$$

Note that this makes sense since we know $F(b_i^1) - F(a_i^1)$. Then again, let $\{(a_i^2, b_i^2)\}$ be the intervals contiguous to E_2 in $[0, 1]$. Using Proposition 2.1 we can now compute $F(b) - F(a)$ for all $[a, b]$ disjoint from E_2 (letting $E = E_1 \cap [a, b]$ there), by the formula

$$F(b) - F(a) = \int_E f(x) dx + \sum_{[a,b]} [F(b_i^1) - F(a_i^1)].$$

Then we compute $F(b_i^2) - F(a_i^2)$ by taking limits as before.

We proceed this way by transfinite induction, taking intersections at limit ordinals.

Of course the concept of arbitrary (countable) ordinals is essential in Denjoy totalization, and Denjoy as well as Lebesgue considered that a major application of the ideas of Cantor's transfinite set theory to analysis and therefore, in their view, an important justification of this theory. (Denjoy has, of course, repeatedly emphasized the relevance of transfinite set theory in analysis (see [D1] and [D2]).

Although Denjoy totalization employs only countably many ordinals for each given derivative f , Denjoy constructed examples of f 's for which his process takes arbitrarily long countably many steps. In other words, the totality of all countable ordinals is necessary in his totalization.

This problem was therefore raised: to what extent is the transfinite induction and the use of the totality of countable ordinals necessary in the problem of reconstructing the primitive (as opposed to its use in a particular process for doing that). See, for example, [P, pp. 170–171], which refers to Lusin's Thesis 1915 as one of the first places where this question has been discussed. Other definitions of integrals (avoiding any use of ordinals) have been proposed, such as the Perron integral, the Kurzweil-Henstock integral, etc. (see [B]). These can be used to recover the primitive of any derivative, but they are hardly constructive in any sense.

3. The complexity of antidifferentiation. The preceding discussion brings us to some recent results which address these problems from the point of view of logic and definability theory—more particularly, in this case, *descriptive set theory*. The idea is to classify the complexity of the operation of antidifferentiation $(*)f \mapsto \int f$. If one can show that this is sufficiently complex (not Borel as we explain in a moment), then this indicates that there is no simple constructive notion of integral (some kind of “super” Lebesgue integration) which is sufficient to invert any derivative. Thus, this shows the necessity of the use of transfinite induction over all the countable ordinals in any constructive such process.

First we have to make precise in what sense we will express the complexity of $(*)$, and this amounts to making clear in what sense a derivative f is considered

as “given.” Since every derivative is a Baire class 1 function, the most reasonable way to consider a derivative as “given” is via a “code” of it as a Baire class 1 function, i.e., in terms of a sequence of continuous functions pointwise converging to it. (One could argue that giving such a sequence gives in some sense too much information about f . But since the point of the results below is that one cannot define simply the primitive of f , *even with this extra information given*, this is even better.)

So let $C[0, 1]$ denote the Polish space of continuous functions on $[0, 1]$ with the uniform metric and let $C[0, 1]^{\mathbb{N}}$ be the Polish space of infinite sequences of continuous functions with the product topology. Let

$$CN = \{\bar{f} \in C[0, 1]^{\mathbb{N}} : \forall x (\{f_n(x)\} \text{ converges})\}$$

be the class of pointwise converging sequences of continuous functions, and for $\bar{f} \in CN$ let $f = \lim \bar{f}$ be defined by $f(x) = \lim f_n(x)$ for all $x \in [0, 1]$. It is not hard to verify that CN is a complete \prod_1^1 subset of $C[0, 1]^{\mathbb{N}}$, so it is not Borel.

Now let

$$\bar{\Delta} = \{\bar{f} \in C[0, 1]^{\mathbb{N}} : \bar{f} \in CN \text{ and } \lim \bar{f} \in \Delta\}$$

be the set of codes of derivatives. The first result here was proven by M. Ajtai several years ago and provides an upper bound for the complexity of $\bar{\Delta}$.

THEOREM 3.1 (AJTAI, UNPUBLISHED). *The set $\bar{\Delta}$ is \prod_1^1 .*

The following lower bound completes the classification.

THEOREM 3.2 [K]. *The set $\bar{\Delta}$ is not \sum_1^1 ; i.e., there is no \sum_1^1 set $S \subseteq CN$ such that for $\bar{f} \in CN$, $\bar{f} \in S \Leftrightarrow \bar{f} \in \bar{\Delta}$. The same holds even for $\overline{b_1\Delta} = \{\bar{f} \in CN : \lim \bar{f} \text{ is a derivative of absolute value } \leq 1\}$.*

This computes the complexity of the domain of the operation of antidifferentiation. For the operation itself one first has the following upper bound.

THEOREM 3.3 (AJTAI, UNPUBLISHED). *The operation of antidifferentiation is Δ_1^1 . More precisely there are \sum_1^1 , \prod_1^1 relations $S, P \subseteq C[0, 1]^{\mathbb{N}} \times \mathbb{R}^2$ such that for $\bar{f} \in CN$, with $f = \lim \bar{f} \in \Delta$ and all $x \in [0, 1]$, $y \in \mathbb{R}$, we have*

$$y < \int_0^x f \Leftrightarrow S(\bar{f}, x, y) \Leftrightarrow P(\bar{f}, x, y).$$

Notice that by standard effective descriptive set theory Theorem 3.3 \Rightarrow Theorem 3.1. Ajtai's proof of Theorem 3.1 and 3.3 used nonstandard models and Denjoy totalization (oral communication). An alternative way to prove Theorem 3.3 is to show that the transfinite process in the Denjoy totalization for f terminates at some ordinal recursive in \bar{f} (where $\lim \bar{f} = f$). This approach uses also a crucial boundedness argument due to H. Woodin.

On the other hand, it was recently established that antidifferentiation is not Borel.

THEOREM 3.4 (DOUGHERTY-KECHRIS [DK]). *The operation of antidifferentiation is not Borel. More precisely, there is no Borel set $B \subseteq C[0, 1]^{\mathbb{N}}$ such that for $\bar{f} \in CN$, with $f = \lim \bar{f} \in \Delta$, $\bar{f} \in B \Leftrightarrow \int_0^1 f > 0$.*

This shows that the totality of countable ordinals is necessary in any constructive process for recovering the primitive, since no “simple analytical” operation suffices for this purpose. We take here as a necessary characteristic of such an operation that it is Borel (in the codes). This is clearly the case for Riemann or Lebesgue integration, taking of limits of sequences, summation of series, etc. So one can argue that arbitrary countable ordinals are intrinsically connected with the operation of recovering the primitive itself, and not only with a particular process of reconstruction, like Denjoy’s.

4. New characterizations of the hyperarithmetical reals. What is behind the preceding results is actually a new characterization of the hyperarithmetical reals. A sequence $\bar{f} \in C[0, 1]^{\mathbb{N}}$ is called *recursive* if it is a recursive sequence of recursive functions. Then Theorem 3.3 can be rephrased as follows.

THEOREM 4.1 (AJTAI, UNPUBLISHED). *Let $\bar{f} \in C[0, 1]^{\mathbb{N}}$ be recursive, $\bar{f} \in CN$, $f = \lim \bar{f} \in \Delta$. Then the primitive $\int f$ is Δ_1^1 . (Similarly relativized to any given real.)*

Note that $F = \int f$ is by its definition a \prod_1^1 singleton (if $f = \lim \bar{f}$, \bar{f} recursive). That it is actually Δ_1^1 is based ultimately on the Denjoy totalization, so this gives a definability consequence of this process, which clearly quantifies its constructive aspects.

Recall that a real $x \in \mathbb{R}$ is *hyperarithmetical* (or Δ_1^1) if $\{r \in \mathbb{Q} : r < x\}$ is hyperarithmetical. Call also a derivative $f \in \Delta$ *recursive* if there is recursive $\bar{f} \in CN$, with $f = \lim \bar{f}$. Then one has immediately from Theorem 4.1 that for a recursive derivative f , $\int_0^1 f$ is a hyperarithmetical real (and similarly relativized to any given real). The main result of [DK] is the converse to this.

THEOREM 4.2 (DOUGHERTY-KECHRIS [DK]). *Let $x \in \mathbb{R}$. Then the following are equivalent:*

- (i) x is hyperarithmetical,
 - (ii) $x = \int_0^1 f$, for some recursive derivative f .
- (Similarly relativized to any given real.)

The proof of this theorem involves effective transfinite induction based on the Recursion Theorem.

One can further combine Theorem 4.2 with Matyasevich’s Theorem to obtain formulas and an equivalent characterization of the hyperarithmetical reals involving only classical notions of analysis.

THEOREM 4.3 (DOUGHERTY-KECHRIS [DK]). *The hyperarithmetic reals are exactly those of the form $\int_0^1 f$, where $f(x)$ is a derivative given by*

$$f(x) = \sum_{n=0}^{\infty} a_n \frac{\max\left(0, 1 - \left| \left(n - [\sqrt{n}]^2 + 1 \right) x - \left([\sqrt{n}] - ([\sqrt{n}]^2) \right) \right| \right)}{n - [\sqrt{n}]^2 + 1},$$

with

$$a_n = (-1)^n \sum_{0 \leq m_1, \dots, m_N \leq 2^{2^N}} \max(0, 1 - Q(n, m_1 \cdots m_N)^2),$$

and Q an exponential polynomial with coefficients in \mathbf{Z} .

Another version of this kind of result can be stated as follows: call a function *analytically expressible* if it can be expressed by an explicit formula involving elementary functions and $\sum_{n=0}^{\infty}$. Then one can show that the hyperarithmetic reals are exactly the reals of the form $\int_0^1 \varphi$, where φ is an analytically expressible derivative. This demonstrates clearly the “definability gap” between a function and its derivative. One can have a derivative given by a simple analytical formula, while its primitive is immensely complex.

5. Some open problems. (A) The first problem is related to definability aspects of so-called “descriptive definitions of integrals” (see [S, Chapters VII, VIII]). These are essentially implicit definitions like the original one of the primitive. For example, the Lebesgue integral F of an integrable function f can be defined as the unique (up to a constant) F which is such that (i) F is absolutely continuous, and (ii) $F'(x) = f(x)$ for almost all x . By replacing, in (i), absolute continuity by more general conditions, one can obtain descriptive definitions of integrals inverting any derivative. The question is whether these conditions can possibly be Borel. (Note that (ii) is clearly Borel and so is the condition of absolute continuity.) This leads to the following

PROBLEM. Is there a Borel relation $B \subseteq C[0, 1]^{\mathbf{N}} \times C[0, 1]$ such that if $\bar{f} \in C^{\mathbf{N}}$, $f = \lim \bar{f} \in \Delta$, and $F \in C[0, 1]$, then $F = \int f \Leftrightarrow (\bar{f}, F) \in B$. We conjecture that the answer is no. (This would be stronger than Theorem 3.4.)

(B) An interesting but a bit vague problem is to come up with simpler formulas in Theorem 4.3. Another more precise question is the following. If in Theorem 4.3 we expand each summand of f in a Fourier series, then we obtain a formula for the hyperarithmetic reals of the form $\int_0^1 \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{nm} \cos m(x - b_n)$, with explicitly given a_{nm} , b_n . Can we express them simply as

$$\int_0^1 \sum_{n=0}^{\infty} a_n \cos n(x - b_n),$$

with recursive $\{a_n\}$, $\{b_n\}$?

(C) The final problem is related to another important result of Denjoy. If a 2π -periodic function $f(x)$ has a trigonometric expansion

$$f(x) = \sum_{n=0}^{\infty} (a_n \cos nx + b_n \sin nx),$$

then the coefficients $\{a_n\}, \{b_n\}$ are uniquely determined by Cantor's Uniqueness Theorem. If f is Lebesgue integrable, the a_n, b_n can be computed by Fourier's formula. But how does one find a_n, b_n from f in the general case? This problem was solved by Denjoy in 1921 (see [D1]) by an extremely complicated procedure, again involving transfinite induction. From the definability point of view this leads to the following two questions:

PROBLEM. Classify the complexity of $T = \{\bar{f} \in CN: \lim \bar{f} \text{ admits a trigonometric expansion}\}$.

PROBLEM. Classify the complexity of the operation $\bar{f} \mapsto \{a_n\}, \{b_n\}$, where $\bar{f} \in T, \lim \bar{f} = f = \sum (a_n \cos nx + b_n \sin nx)$.

From its definition T is Σ_2^1 and this operation is Δ_2^1 on T . However, it appears that Denjoy's constructive process for recovering a_n, b_n ought to lead to a substantial lowering of this complexity, perhaps at the level of Σ_1^1 or Π_1^1 nonmonotone inductive definitions. If this is correct and if one can show corresponding lower bounds, this would lead to a classification of a natural and basic concept of analysis which falls between two levels of the projective hierarchy. This certainly would be a very interesting phenomenon. Also, if it is indeed true that the complexity of computing the trigonometric expansion is quite high above Borel, this would give a nice definability "explanation" of the considerable difficulty of Denjoy's procedure. At this point, however, this is only speculation, since it is not even known whether $\bar{f} \mapsto \{a_n\}, \{b_n\}$ is Borel or not.

REFERENCES

- [B] P. Bullen, *Non-absolute integrals: A survey*, Real Anal. Exchange, **5** (1979–80), 195–259.
- [D1] A. Denjoy, *Leçons sur le calcul des coefficients d'une série trigonométrique*, I–IV, Gauthier-Villars, Paris, 1941–49.
- [D2] —, *L'énumération transfinie*, Gauthier-Villars, Paris, 1946.
- [DK] R. Dougherty and A. Kechris, *The complexity of antidifferentiation*, Nov. 1985; Addendum, *Analytical expressions for the hyperarithmetical reals*, Feb. 1986; mimeographed circulated notes.
- [K] A. Kechris, *Classifying the complexity of the set of derivatives*, July 1985, mimeographed circulated notes.
- [L1] H. Lebesgue, *Leçons sur l'intégration et la recherche des fonctions primitives*, Gauthier-Villars, Paris, 1904.
- [L2] —, *ibid.*, 2eme éd., Paris, Gauthier-Villars, 1928.
- [P] I. Pesin, *Classical and modern integration theories*, Academic Press, New York, 1970.
- [S] S. Saks, *Theory of the integral*, 2nd rev. ed., Dover, 1964.

CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA 91125, USA